ARTIFICIAL NEURAL NETWORK COMPARISON OF EXPERT AND NOVICE PROBLEM-SOLVING STRATEGIES

Ronald H. Stevens, Ph.D.,
Alina C. Lopo, M.D., Ph.D.
Department of Microbiology and Immunology
UCLA School of Medicine
Los Angeles, California 90024-1847

The successful strategies of second-year medical students were electronically captured from computer-based simulations in immunology and infectious disease and were used to train artificial neural networks for the rapid classification of subsequent students' and experts' strategies on these problems. Such networks could categorize problem solutions of other students as successful or non-successful >85% of the time. These neural networks, however, performed poorly (as low as 13%) when classifying experienced immunologists' or internists' successful performances, suggesting an ability to distinguish between novice and expert strategies. The neural networks also identified a group of students who framed the infectious disease problems correctly, but had difficulty discriminating between differential diagnoses.

INTRODUCTION

We have been exploring the ability of artificial neural networks to classify the performances of medical students who are engaged in problem solving in multiple disciplines and have shown that neural networks trained with the successful problem performances of students can accurately recognize the strategies of new students on these same problems > 85% of the time [1].

This success rate would not be unusual were separate neural networks trained for each problem in each discipline and if subsequent students performances were evaluated by these individual neural networks. In fact however, within each discipline, a single neural network has encapsulated successful student strategies across the 6-7 different problems composing the problem set. This suggest two distinct abilities of the trained neural networks, the ability to recognize successful strategies from problem performances, and the ability to discriminate each problem from the others in the set, even when they are quite close conceptually.

These broad capabilities suggested that neural networks trained with students' performances could provide revealing information not only about the nature of students approaches to the problems, but also about the nuances constituting various levels of expertise. To pursue these studies we have used two very different problem spaces, immunology and infectious disease.

The IMMEX::IMMUNOLOGY problem set is basic-science-oriented and students need an understanding of molecular immunology as well as knowledge of the principles of flow cytometry, RFLP, gel retardation etc. Studies of student performances on these problems reveal considerable search and the generation and discarding of alternative hypotheses. By contrast, the IMMEX::INFECTIOUS DISEASE problems are clinical and diagnostic in scope.

In this study we wanted first to determine how suitably trained neural networks would classify new students performances in these two different problem domains to determine sources of predictive error, and then to determine if such neural networks could perhaps distinguish expert from novice performances.

METHODS

The IMMEX Problem-Solving Format

The approach is based on the cognitive principles of having a starting condition (i.e. Case History), a goal condition (i.e., Diagnosis) and the access to the information needed to transit these conditions. Each problem starts with a patient history which contains sufficient information for the generation of hypotheses regarding the possible immune defect or in the case of the infectious disease problems, a process and infectious agent involved. Students performing these problems then access additional information and laboratory tests from 50-70 different menu items which can be used to verify/reject hypotheses. When they are confident of the patient's immune defect/disease process, a diagnosis can be made. The details of the software and its implementation have been described in detail [2].

During the problem solving, a transaction database records the student's selection of information, time, score, diagnosis, etc. This can be accessed by search-path mapping software which displays the students sequential requests for more information and can therefore reconstruct individual or group problem solving performances [3]. The IMMEX::ANALYSIS software also saves the test selections returned from queries made to the database and prepares them for insertion into the IMMEX::NEURAL software for generation of artificial neural networks and the classification of subsequent performances.

Construction and Training of Artificial Neural Networks with Student Performances

Multi-layer backpropagation neural networks were trained using over 400 performances of students who solved one or more of seven different problems in immunology or six problems in infectious disease. The training data for the back propagation neural networks [4] were from individual student problem solving performances which had been collected under conditions requiring students working on their own.

The input data for the neural network was derived as follows. As students progress through the problems the sequence of their test selections was recorded in the form of two-test classifying characteristics. For instance in Figure 1, the classifying characteristics would be "Start To FACS CD4/CD3", " FACS CD4/CD3 TO T-Cell Proliferation", "T Cell Proliferation TO MHC mRNA" etc. Each unique classifying characteristic of the training set constitutes an input node. In Immunology the neural network constructed consisted of 533 input neurons (one for each classifying characteristic), 40 hidden neurons, and output neurons, one for each problem in the problem set. These layers were fully interconnected by weighted links; the momentum was 0.9, the learning rate was 0.06 and the network was trained to a 0.005 sum of errors [1]. During testing, a student's test selection is presented to the neural network and the problem-specific output weights collected. The problem-specific output weights range from 0 to 1. This process is repeated for each test selection made by the student until the completion of the problem, resulting in a series of output weights for each problem which can be displayed as histograms. Successful performances are indicated by high output weights for the relevant problem and low for the other problems (Figure 1). Unsuccessful, and false negative performances are often indicated by low output weights across all problems.

For the infectious disease problems, the number of input classifying characteristics was 654, the number of hidden nodes was 20, and there were six problem specific output neurons. The learning rate and momentum were as for the immunology problems.

Collection of Expert Immunologists and Infectious Disease Internist's Performances

Two different approaches were used to collect expert problem performances. In Immunology, anonymous problem performances by experienced immunologists were collected over a 3-day period at the 1993 American Association of Immunologists meeting. Seventy-six percent of the participants held the title Assistant Professor or higher, providing an indication of the level of expertise. Of the problems completed by experts, there were 123 performances where the problem was solved and 55 instances where the diagnosis was missed. This frequency of solutions (69%) was slightly higher than that of UCLA second-year medical students under testing conditions (302/450 or 67%) this past year. While this expert performance value may seem low, consideration should be given to the fact that each student had over 10 hr. practice on similar problems and their performance during the exam would account for 50% of their total grade. Of these 178 immunologists' performances, 87 were on problems where student performances were used to train artificial neural networks. These performances constitute the testing data for this part of the study.

Expert performances on the infectious disease problems were collected from individuals holding the position of Chief Resident or higher in the Department of Medicine at UCLA or its affiliated hospitals. The frequency of solutions of the internists (81%) was slightly lower than that of UCLA second-year medical students (87%), again most likely for the reasons mentioned earlier.

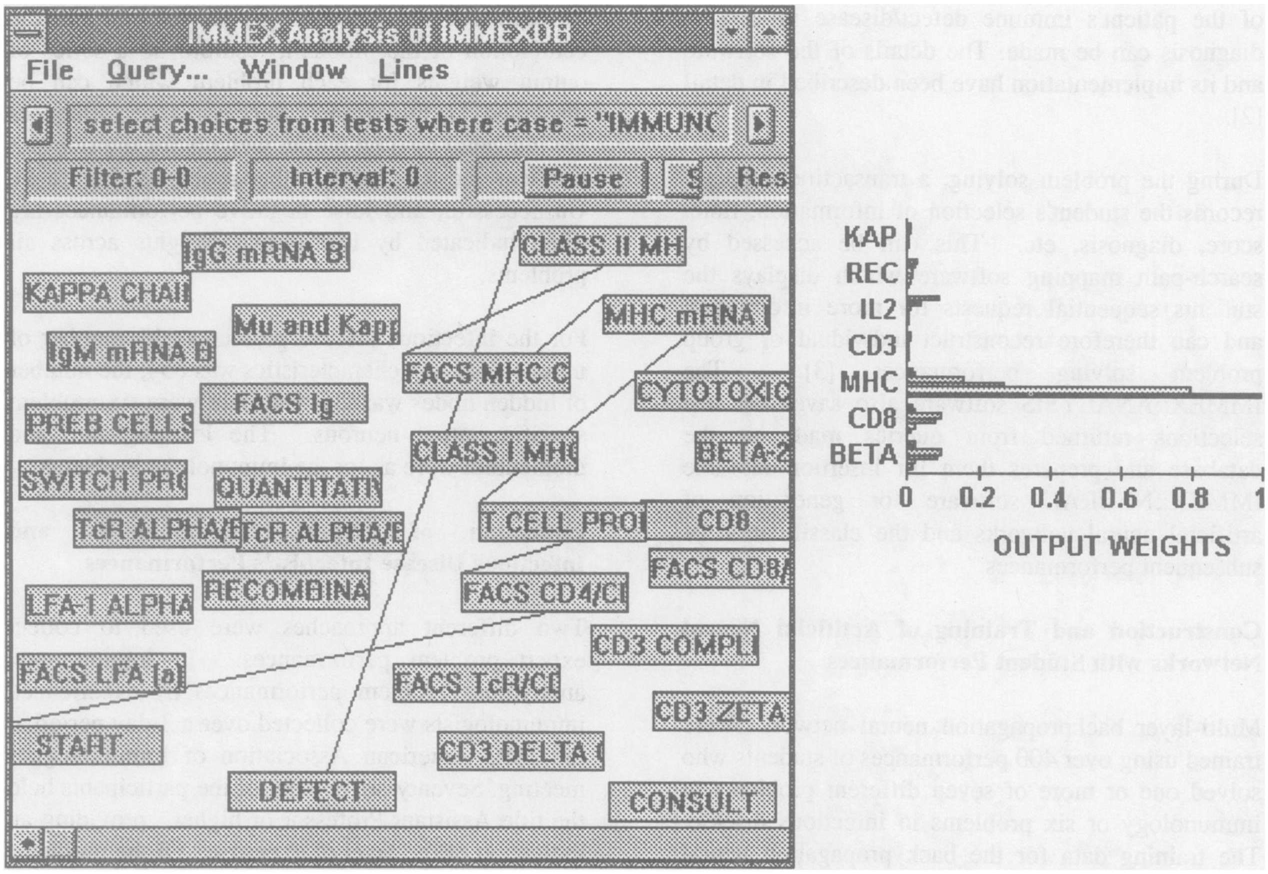

Figure 1 A Comparison of IMMEX::ANALYSIS search path mapping and IMMEX::NEURAL which provides an interpretation of the analysis output. These figures follow the progression of one student as tests were selected during solving an immunology problem. The lines connecting the boxes show the sequence of a student's tests. The histograms show the output weights returned from a trained neural network as each of these test selections were presented to the trained neural network.

RESULTS

A. Expert Immunologists' Performances

True negative performances (where the correct diagnosis was not obtained) of both the students (47/49) and immunologists (20/21) were accurately detected by the artificial neural networks trained with student performances (Table 1). The neural networks identified 33/44 (75%) true positive performances (where the diagnosis was made) of second-year UCLA medical students and 18/26 (69%) true positive performances of first-year George Washington University medical students, all of which were obtained under testing conditions. Thus subsequent student performances when presented to the trained neural networks, were correctly classified as having solved or not solved a particular problem >85% of the time. In contrast, only 23/66 (35%) of the immunologists' performances were identified by the student trained neural networks. The true positive immunologists' performances which were detected by the neural networks were not uniform across the problems but ranged from a high of 75% to a low of 13% (Table 1). The above results indicate that the sequence of actions employed by immunologists in solving the same problems are not well encapsulated by neural networks trained on students' successful problem performances.

TABLE 1

		STUDENTS' PERFORMANCE TRUE +	TRUE -
NETWORK CLASSIFICATION	+	46	2
	-	17	47

		IMMUNOLOGISTS' PERFORMANCE TRUE +	TRUE -
NETWORK CLASSIFICATION	+	23	1
	-	43	20

Table 1. Contingency Tables of Student and Immunologists Problem Performances as Classified by Student-Trained Artificial Neural Networks trained on IMMEX Immunology problems.
The sensitivity for the individual problems performed by the immunologists was: Bare Lymphocyte Syndrome (27%), CD3 Complex Deficiency (13%), Beta-2 Microglobulin Defect (25%), Recombinase Defect (27%), and IL-2 Promoter Defect (75%). Varying the neural network output weight decision threshold values between 0.45 and 0.65 did not produce significant differences in the above classifications. The student and immunologist distribution of true positive and false negative were significantly different (Pearson $\chi^2 = 18.46$ $P<0.0005$).

B. Infectious Disease Problems

As discussed earlier, the infectious disease simulations differ significantly from the immunology problems in that they more closely parallel the clinical diagnostic process. As such, we were as much interested in the sensitivity and specificity performance of the student trained artificial neural networks as we were in how well they would discriminate between expert and novice performances. Forty-three percent (62/144) of the student true negative performances were classified by the infectious disease trained neural network as positive. This was in direct contrast to the immunology performances and suggested an appropriate strategy was being used but an incorrect diagnosis was resulting. Visual analysis of these student performances by search path mapping using IMMEX::ANALYSIS confirmed the inability of these students to clearly distinguish between related differential diagnoses. The infectious disease internists did not have this difficulty (Table 2).

Similar to the immunology problems, the student-trained neural networks identified 176/253 (70%) of true positive performances for medical students. These same neural networks identified internists performances 61% of the time. As with the immunology basic science problems, the true positive internists' performances detected by the neural networks were not uniform across the problems but ranged from a high of 100% to a low of 17%.

CONCLUSIONS

Artificial neural networks have had a broad applicability in medical decision making [5]. Our studies extend these efforts to medical education and indicate that appropriately trained artificial neural networks may be useful tools which can be used not only for routine (and rapid) evaluation of student problem solving performances, but also which may be used to discriminate between novice and expert performances, particularly on the more difficult problems.

We are currently acquiring a sufficient number of immunologists' and infectious disease experts' problem performances to train "expert" artificial neural networks. With these networks, in an evaluation setting, student "passing" may consist not only of solving a series of problems, but by solving them with a strategy better represented in the expert neural network rather than the novice-trained neural network.

TABLE 2

		STUDENTS' PERFORMANCE	
		TRUE +	**TRUE -**
NETWORK CLASSIFICATION	+	176	62
	-	77	82

		ID INTERNISTS' PERFORMANCE	
		TRUE +	**TRUE -**
NETWORK CLASSIFICATION	+	33	3
	-	21	12

Table 1. Contingency Tables of Students' and Infectious Disease Experts' Problem Performances as Classified by Student-Trained Artificial Neural Networks.

The sensitivity for the individual problems performed by the ID experts was: Bacterial endocarditis (57%), Rheumatic Fever (100%), Listeria (57%), Disseminated M. tuberculosis (83%), M. avium intracellularae (17%) and Salmonella osteomyelitis (83%). Varying the neural network output weight decision threshold values between 0.45 and 0.65 did not produce significant differences in the above classifications.

REFERENCES

[1]. Stevens RH, Najafi K. Artificial neural networks as adjuncts for assessing medical students' problem solving performances on computer based simulations. *Computers & Biomedical Research* 1993; 26:172-187.

[2]. Stevens RH, McCoy JM, Kwak AR. Solving the problem of how medical students solve problems. *MD Computing.* 1991; 8:13-20.

[3]. Kwak AR, Stevens RH. Administering a microcomputer-based problem-solving examination. *Journal of Biological Computing* 1990;17(3):9-13.

[4]. Rumelhart D, Hinton GE, McClelland JL. *Parallel Distributed Processing,* 1, Ch. 2, The MIT Press, 1986.

[5]. Reggie, J. Neural Computation in Medicine. *Artificial Intelligence in Medicine* 1993; 5(2):143-158.

ACKNOWLEDGMENTS

Supported in part by funding from the United States Public Health Services, Bureau of Health Professions. Special thanks to Drs. Phyllis Kind, Barbara Dau and Ann Linton at GWU for providing student performances, to the immunologists who participated in the IMMEX Problem-Solving Software Workshop at the 1993 AAI meeting in Denver, and to the infectious disease experts at UCLA, the UCLA Affiliated Hospitals, and the Health Sciences Center, University of Texas at San Antoino